WUS ZUSAMMENFASSUNG

SIMONE GUGGIARI
ETH ZÜRICH 2015
LEGI # 14-932-867

WAHRSCHEINLICHKEITSTHEORIE (WUS)

Grundlegende Kombinatorik

Die folgende Tabelle bezieht sich auf das Urnenmodell (»Variationen«)

ziehe k aus n	geordnet	ungeordnet
mit zurücklegen	n^k	$\binom{n+k-1}{k}$
ohne (repetition)	$\frac{n!}{(n-k)!}$	$\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$ → bin. coefficient (combinations)

$n!$ = permutations

Ereignisraum Ω

(collection of all results) – sample space

Menge aller möglichen Elementarereignisse ω des betrachteten zufallsexperiments

es. Würfel: $\Omega = \{1,2,3,4,5,6\}$

Lanci 2 monete: $\Omega_2 = \{(H,H),(H,T),(T,H),(T,T)\}$

Ereignis A

(evento) – una possibilità

Teilmenge des Ergebnisraums Ω ($A \subseteq \Omega$), also eine Kollektion aus Elementarereignissen. Kann auch mehrere Ergebnisse des zufallsexperiment umfassen (raminfee)

es. Gerader Würfelwurf: $A = \{2,4,6\} \subseteq \Omega$

Beobachtbare Ereignisse $\mathcal{F}$

(power set) – tutti gli eventi possibili

Ereignisse, die man wirklich beobachten, also messen kann.

→ $\mathcal{F} \subseteq 2^\Omega = \mathcal{P}(\Omega) = \{A \mid A \subseteq \Omega\}$

Normalerweise (Ω endlich oder abzählbar), wählt man $\mathcal{F} = 2^\Omega$, also alle möglichen Ereignisse des Ereignisraums. (tutti i gruppi / le combinazioni)

$\mathcal{F} \subseteq 2^\Omega$ σ-Algebra if
- $\Omega \in \mathcal{F}$
- $\forall A \in \mathcal{F} \Rightarrow \bar{A} \in \mathcal{F}$
- $\forall (A_m)_{m \in \mathbb{N}_0} = \{A_0, A_1, A_2, \dots\} \subset \mathcal{F} \Rightarrow \bigcup_{i=0}^{\infty} A_i \in \mathcal{F}$

Mengenoperationen auf Ereignisse

die funktionieren genau so, wie man es sich auch intuitiv vorstellen würde:

- Schnittmenge $A \cap B$: ereignis, dass gleichzeitig A und B eintreten (si verificano)
- Vereinigung $A \cup B$: ereignis, dass A oder B (oder auch beide) eintreten
- Komplement $\bar{A} = A^c$: ereignis, dass A nicht eintritt

Wahrscheinlichkeitsmass P

Funktion, welche einem Ereignis einen Wert in $[0,1]$ zuordnet, also eine abbildung $P: \mathcal{F} \to [0,1]$, $A \mapsto P[A]$

Axiome & Rechnenregeln

A1) $\forall A \in \mathcal{F}$: $P[A] \geq 0$

A2) $P[\Omega] = 1$

A3) if A_i paarweise disjunct ($\forall i \neq j$ $A_i \cap A_j = \emptyset$): $P[\bigcup A_i] = \sum P[A_i]$ → $P[A \cup B] = P[A] + P[B]$

(·) → il punto significa "paarweise disjunct"

R1) $P[\bar{A}] = 1 - P[A]$ — $\Omega = A \cup \bar{A}$ ($\Omega = A \cup \bar{A}$, $A \cap \bar{A} = \emptyset$) → $1 = P[\Omega] = P[A] + P[\bar{A}]$

R2) $P[\emptyset] = 0$

R3) $A \subseteq B \Rightarrow P[A] \leq P[B]$ — $B \cap A^c = B \setminus A$ — $B = (B \cap A) \cup (B \cap \bar{A}) = (A \cap B) \cup (B \setminus A)$ — $P[B] = P[A] + P[B \setminus A] \geq P[A]$

R4) $P[A \cup B] = P[A] + P[B] - P[A \cap B]$ (Additionsregel)

R5) $P[\bigcup_{i=1}^n A_i] \leq \sum_{i=1}^n P[A_i]$ (not disjunct) – union bound

R6) $P[\bigcup_{i=1}^n A_i] = \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P[A_{i_1} \cap \dots \cap A_{i_k}]$

Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$

es charakterisiert ein Zufallsexperiment und umfasst einen Grundraum Ω, die beobachtbaren Ereignisse $\mathcal{F}$ und eine Wahrscheinlichkeitsabbildung P.

Laplace-Raum

(gleichverteilung)

es ist ein endlicher Wahrscheinlichkeitsraum, in welchem alle Elementarereignisse gleich wahrscheinlich sind. Das wahrscheinlichkeitsmass P in diesem raum heisst diskrete Gleichverteilung. Für alle ereignisse $A \subseteq \Omega$ gilt

$P[A] = \frac{|A|}{|\Omega|}$ Frequentistisch: $P[A] = \lim_{n \to \infty} \frac{1}{n} \#\{i \in \{1,\dots,n\} \mid A \text{ happens}\}$

Bedingte wahrscheinlichkeit

(probabilità condizionale)

Die bedingte wahrsc. von B gegeben A entspricht der Wahrscheinlichkeit, dass B eintritt, wenn man schon weiss, dass A eingetreten ist.

$P[B|A] = \frac{P[A \cap B]}{P[A]}$ multiplication rule: $P[A \cap B] = P[A] \cdot P[B|A] = P[B] \cdot P[A|B]$

Die pfadregel, nach der Wahrscheinlichkeiten in einem Wahrscheinlichkeitsbaum multipliziert werden, um die Wahrsch. eines "Blattes" zu erhalten, entspricht eine verkettung bedingter Wahrscheinlichkeiten

Satz der totalen Wahrscheinlichkeit

$\triangleright$ A_i disjunkte zerlegung von Ω ($\bigcup_{i=1}^n A_i = \Omega$).
Für beliebige Ereignisse B gilt dann:

$P[B] = \sum_{i=1}^n P[A_i \cap B] = \sum_{i=1}^n P[A_i] \cdot P[B|A_i]$

der satz bedeutet gerade, dass man die Wahrscheinlichkeit eines ereignisses "ermitteln" in einem Wahrscheinlichkeitsbaum bekommen kann, wenn man die einzelnen Pfade addiert, die zu ihm führen.

Satz von Bayes

mit ihm lassen sich bedingte wahrscheinlichkeiten "umkehren" $\triangleright$ A_i: disj. zerlegung von Ω ($\bigcup A_i = \Omega$), und B ein weiteres ereignis. dann gilt $\forall k$

$P[A_k|B] = \frac{P[A_k \cap B]}{P[B]} = \frac{P[A_k] \cdot P[B|A_k]}{\sum_i P[A_i] \cdot P[B|A_i]}$ ← bei wahrscheinlichkeit

$P[B|A] = \frac{P[A|B] \cdot P[B]}{P[A]}$

Unabhängigkeit

zwei ereignisse A,B heissen unabhängig falls $\Rightarrow P[A \cap B] = P[A] \cdot P[B]$

assuming ($P[A] \neq 0, P[B] \neq 0$)

A,B unabhängig $\Leftrightarrow P[B|A] = P[B] \Leftrightarrow P[A|B] = P[A]$

Anschaulich bedeutet Unabhängigkeit, dass das eintreten des einen Ereignisses keinen influss auf die Wahrsch. des anderen hat. Deshalb ist auch die Kovarianz $Cov(X,Y)$ zweier Zufallsvariablen X und Y gleich null, falls sie unabhängig sind (⚠, umkehrung gilt nicht!)

paarweise unabh. ↗

[unabh $\Rightarrow$ Cov = 0 ($\not\Leftarrow$)

Die ereignisse $A_i, i \in \{1,\dots,n\}$ heissen unabhängig, wenn für jede Teilfamilie T von A_i gilt:

$\forall\, 1 \leq j_1 < \dots < j_k \leq n$, $T = A_{j_1} \cap \dots \cap A_{j_k}$ $\quad P[T] = \prod_{m=1}^{k} P[A_{j_m}]$

⚠ Paarweise unabhängenheit bedeutet nicht, dass alle elemente unabhängig sind.

Gegenbeispiel: würfelwurf: A = erste zahl gerade, B = zweite —, C = beide gerade oder beide ungerade

Zufallsvariablen (ZV)

Zufallsvariable X

es ist eine Funktion $X: \Omega \to \mathbb{R}$, welche jedem elementarereignis eine zahl zuordnet, welche repräsentiert, wofür wir uns bei unserem zufallsexperiment überhaupt interessieren

es:
- Augenzahl beim würfelwurf
- Anzahl köpfe bei k-maligem Münzwurf
- Gewicht einer aus einem sack gezogenen kartoffeln

Eine zufallsvariable mit abzählbarem wertebereich heisst diskret, ansonsten ist sie stetig.

Indikatorfunktion $\mathbb{1}$

ist eine spezielle Zufallsvariable, welche angibt, ob ein Ereignis A eingetreten ist (si è verificato)

$\mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$

Gewichtsfunktion

(diskret)

die Gewichtsfunktion $f_X(t): W(X) \to [0,1]$ ordnet jeden wert t, den eine zufallsvariable annehmen kann, seine Wahrscheinlichkeit zu. Viel intuitiver aber auch erlaubt ist die schreibweise $P[X=t]$, wobei $t \in W(X)$ (interessa)

Anschaulich gibt $P[X=t]$ auf die frage "Auskunft "mit welcher wahrsch. nimmt zufallsvariable X den wert t ?"

⚠ die Gewichtsfunktion ist nur für diskrete Zufallsv. sinnvoll (denn bei einer stetigen verteilung hat ein einzelnereignis immer wahrsch. = 0)

Dichtefunktion

(stetig)

die dichtefunktion ist das Analog zur gewichtsfunktion für stetige zufallsvariablen. f_X. Sie gibt den Änderung ihrer verteilungsfunktionen an, und charakterisiert sie so | vollständig. ⚠ non è il singolo punto che ha un valore

eigenschaften: $\forall t \in W(X): f_X(t) \geq 0$ ($\not\Leftarrow$)

$\int_{-\infty}^{\infty} f_X(t)\,dt = 1$ (area = 1)

Transformation

$\triangle$ X zufallsv. mit dichtefunkt. f_X und $Y = g(X)$ bijektiv. ($X = g^{-1}(Y)$)
Dann hat Y die dichtefunkt:

$f_Y(t) = \left|\frac{d}{dt} g^{-1}(t)\right| \cdot f_X(g^{-1}(t))$

es: $Y = aX + b$. then $f_Y(t) = \frac{1}{a} \cdot f_X\left(\frac{t-b}{a}\right)$

Verteilung μ

(boreich)

Die verteilung $\mu_X: \mathcal{B} \to [0,1]$ mit $\mathcal{B} \subseteq W(X)$ ist das analog zu gewichts/dichtefunktion für Bereiche (intervalle/regione) = Teilmengen der Wertemenge einer Zufallsvariablen X

Anschaulich beantwortet $\mu_X(B)$ wie gross die Wahrsch. ist, dass X einen Wert aus B annimt.

Bei stetigen Zufallsvariablen ist B oft Teilintervall von $W(X)$

Verteilungsfunktion F

→ derive for f_X

die Verteilungsfkt. $F_X : \mathcal{W}(X) \to [0,1]$ gibt über die Wahrsch. Auskunft (info), mit welcher eine Zufallsvariable X höchstens einen bestimmten Wert erreicht.
Interhuer und erlaubt ist die Schreibweise $P[X \le t]$, $t \in \mathcal{W}(X)$

eigenschaften:
- F_X monoton wachsend ↗, rechtsstetig
- $\lim_{t \to -\infty} F_X(t) = 0$ · $\lim_{t \to \infty} F_X(t) = 1$
- if diskret: $F_X(t) = \sum_{\substack{i \in \mathcal{W}(X) \\ i \le t}} p_X(i) = \sum_{\substack{i \in \mathcal{W}(X) \\ i \le t}} P[X = i]$
- if stetig: $F_X(t) = \int_{-\infty}^{t} f_X(u)\,du$

- for disk/stetig gilt $P[a < X \le b] = F_X(b) - F_X(a)$
  if stetig, you can switch $<, \le$ and $>, \ge$ (every single value has prob = 0)

Quantil und median

für $\alpha \in]0,1[$ heisst $q_\alpha = F_X^{-1}(\alpha)$ das α-Quantil der Verteilungsfunk. F.
$\Rightarrow F_X(q_\alpha) = \alpha$. Es bedeutet, dass die Zufallsvariable X mit wahrsch. α Werte zwischen $]-\infty, q_\alpha]$ annimmt.

Das $\frac{1}{2}$-Quantil eine Verteilung heisst Median.

q_α divide le prob. in α e $(1-\alpha)$

Erwartungswert E → constant

er zeigt, welchen Wert eine Zufallsvariable bei vielen Wiederholungen des Zufallsexperiment durchschnittlich (in media) annehmen wird.
Oft schreib man auch $E[X] = \mu$ (△ different from Verteilung)

diskret: $E[X] = \sum_{t \in \mathcal{W}(X)} t \cdot P[X = t]$ (media ponderata)

stetig: $E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)\,dx$

erlaubte transformationen:
- $Y = g(X)$ diskret: $E[Y] = \sum_{t \in \mathcal{W}(X)} g(t) \cdot P[X = t]$
- $Y = g(X)$ stetig: $E[Y] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)\,dx$
- $E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y)\,dx\,dy$
- $E[E[X]] = E[X]$
- $Z = g(X) \cdot h(Y)$, X, Y unabhängig: $E[Z] = E[g(X)] \cdot E[h(Y)]$ △ umkehrung nicht. gilt (?)
- $Z = \alpha X + \beta Y$ (linearity) $E[Z] = \alpha E[X] + \beta E[Y]$
- $Y = a + \sum_{i=1}^{n} b_i X_i$ $E[Y] = a + \sum_{i=1}^{n} b_i \cdot E[X_i]$
- $E[\sum X_i] = \sum E[X_i]$ ↑ they all have to be finite
- $E[\prod X_i] = \prod E[X_i]$ (unabhängig/unkorreliert)

Erwartungswert einer Funktion mehrere ZV

△ $X_1, \dots, X_n$ ZV mit gemeinsamer gemeinsamer Dichtefkt ρ / gemeinsamer Verteilfkt f, $Y = g(X_1, \dots, X_n)$. Dann gilt

diskret: $E[Y] = \sum_{x_1, \dots, x_n} g(x_1, \dots, x_n) \cdot p(x_1, \dots, x_n)$

stetig: $E[Y] = \int \dots \int g(x_1, \dots, x_n) \cdot f(x_1, \dots, x_n)\,dx_n \dots dx_1$

Varianz Var $= \sigma^2$

ist ein Mass für die Streuung (dispersione) einer ZV. Sie stellt den übergang (passaggio) zwischen der ZV selbst und ihrer standardabweichung dar (rappresenta). Oft schreibt man $\mathrm{Var}[X] = \sigma^2$

$\mathrm{Var}[X] = E[(X - E[X])^2] = E[Y^2]$, $Y = X - E[X]$ (expanded form)

if $E[X^2] < \infty$ $\mathrm{Var}[X] = E[X^2] - E[X]^2$ (in out)

X diskret: $\mathrm{Var}[X] = \sum_{t \in \mathcal{W}(X)} (t-\mu)^2 \cdot P[X = t]$ ($\mu = E[X]$)

X stetig: $\mathrm{Var}[X] = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f_X(x)\,dx$

for $Z = a + bX$ $\mathrm{Var}[Z] = b^2 \mathrm{Var}[X]$ beweis: use expanded form

for $X_1, \dots, X_n$ unabhängig: oder unkorreliert (paarweise) $\mathrm{Var}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathrm{Var}[X_i]$ △ you need the info! otherwise not computable

general case: $\mathrm{Var}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathrm{Var}[X_i] + 2 \sum_{i<j} \mathrm{Cov}(X_i, X_j)$

$\mathrm{Var}[X_1 + X_2] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2] + 2\mathrm{Cov}(X_1, X_2)$

unabhängig $\Rightarrow$ paarweise unabhängig $\Rightarrow$ unkorreliert
$\not\Leftarrow$ $\not\Leftarrow$
(stronger)

Standardabweichung σ (deviazione standard)

ist das Mass für die Abweichung (deviaz) einer ZV von ihrem Erwartungswert.

$\sigma = \sqrt{\sigma^2} = \sqrt{\mathrm{Var}[X]}$ (nd formulario = s)

Covarianz Cov

ist ein Mass für den Zusammenhang (relaz/rapporto) zwischen zwei ZV X und Y, bzw (rispettivam) ihrer Streuung (dispersione)
→ how much 2 random variables change together >0 → they have similar behavior

$\mathrm{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])] = E[X \cdot Y] - E[X] \cdot E[Y] = \frac{\mathrm{Var}[X+Y] - \mathrm{Var}[X] - \mathrm{Var}[Y]}{2}$

unabh $\Rightarrow$ Cov = 0 $\Rightarrow$ unkorreliert
$\not\Leftarrow$

properties
- $\mathrm{Cov}(X,X) = \mathrm{Var}[X]$
- $\mathrm{Cov}(X, Y+Z) = \mathrm{Cov}(X,Y) + \mathrm{Cov}(X,Z)$
- $\forall a \in \mathbb{R}: \mathrm{Cov}(X,a) = 0$
- $\forall b \in \mathbb{R}: \mathrm{Cov}(X, bY) = b \cdot \mathrm{Cov}(X,Y)$

$\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$

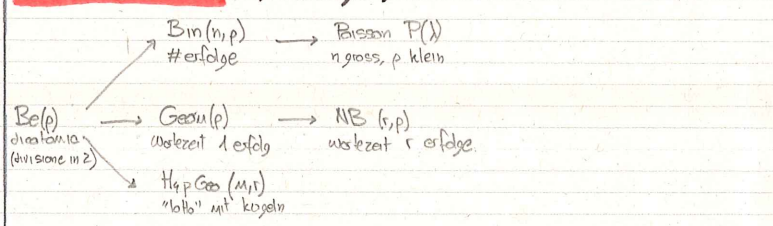
Korrelation ϱ

wie die Kovarianz, ist auch die Korrelation ein Mass für den Zusammenhang zwischen zwei Zufallsvariablen X und Y.
In gegensatz zu dieser ist die Korrelation jedoch dimensions- und einheitslos, und eignet sich (è adatta) deshalb für Vergleiche verschiedener Zusammenhänge (relazioni)

$\varrho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}[X] \cdot \mathrm{Var}[Y]}}$

eigenschaften:
- $\varrho(X,Y) \in [-1,1]$.
- $|\varrho(X,Y)| = 1 \Leftrightarrow \exists a, b \in \mathbb{R}, b \neq 0$ with $P[Y = a + bX] = 1$
  then we say X, Y are "perfekt korreliert"

DISTRIBUTIONS (Verteilungen)



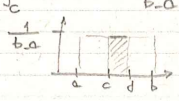
*** forgetfulness** property (Gedächtnislosigkeit)

$P[X \ge t+s \mid X \ge s] = P[X \ge t]$ (diskret)
(stetig—here it's interchangeable)

Die Wahrsch. dass man noch einen weiteren Zeitabschnitt t warten muss, hängt nicht davon ab, wie lange man schon gewartet hat.

*** summe** △ $X_1, \dots, X_r \sim \mathrm{Geom}(p)$ i.i.d. $\Rightarrow X_1 + \dots + X_r \sim \mathrm{NB}(r,p)$

*** interval**: per un intervallo: $P[X \in [c,d]] = \int_c^d f_X(t)\,dt = \int_c^d \frac{1}{b-a}\,dt = \frac{d-c}{b-a}$



*** addit**: $X \sim P(\lambda > 0)$? unabhängig $\Rightarrow X+Y \sim P(\lambda + \mu)$
$Y \sim P(\mu > 0)$

DISKRET

Diskrete Gleichverteilung (=laplace)

name

description

Einfachste Verteilung, bei welche alle n Werte $\in \mathbb{Z}$ zwischen zwei Schranken $a, b \in \mathbb{Z}$ gleich wahrsch. sind.

(math)

es: Ergebnis eines Münz- oder Würfelwurf

es

graph

$b-a+1=n$

schreibweise

$P[X=t]$: $\begin{cases} \frac{1}{n} & \text{if } t \in [a,b] \\ 0 & \text{else} \end{cases}$

$\mu = E[X]$: $\frac{a+b}{2}$

$\sigma^2 = Var[X]$: $\frac{n^2-1}{12}$

$F_X(t) = P[X \le t]$: $\begin{cases} 0 & \text{if } t<a \\ \frac{t-a+1}{n} & \text{if } t \in [a,b] \\ 1 & \text{if } t>b \end{cases}$

$\mu_X(B)$ (bes. Verteilung): $\frac{|B|}{n}$

Bernoulli-Verteilung

Eine Bernoulli-verteilte ZV mit parameter p nimmt genau die beiden Werte 0,1 mit den wahrsch. $(1-p), p$ an.

Also ist sie auch die Indikatorfunktion $\mathbb{1}_A$ für ein Ereignis A mit $P[A]=p$

$\mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$

$X \sim Be(p)$

$P[X=t]$: $\begin{cases} 1-p & \text{if } t=0 \\ p & t=1 \end{cases}$ also wartable $\rightarrow$ $p^t \cdot (1-p)^{1-t}$, $t \in \{0,1\}$

$\mu = E[X]$: p

$\sigma^2 = Var[X]$: $p \cdot (1-p)$

Binomial-verteilung

Die Binomialverteilung beschreibt die Anzahl der Erfolge bei n-unabhängige 0-1 experimenten mit erfolgwahrscheinlichkeit p.

$X = \sum_{i=1}^{n} Y_i, \quad Y_i \sim Be(p)$

es: #Köpfe beim 10-mahligen Münzwurf

$X \sim Bin(n,p)$

$P[X=t]$: $\binom{n}{t} \cdot p^t \cdot (1-p)^{n-t}$

$\mu = E[X]$: $n \cdot p$

$\sigma^2 = Var[X]$: $n \cdot p \cdot (1-p)$

* chernoff-schranken

$F_X(t) = P[X \le t]$: $1 - (1-p)^t$

Geometrische verteilung

Die Geometrische Verteilung mit parameter p beschreibt die wartezeit auf den ersten erfolg, bei einer folge unabhängiger 0-1 experimente mit erfolgwahrscheinlichkeit p.

$X = \inf\{ i \in \mathbb{N} \mid Y_i = 1 \}, \quad Y_i \sim Be(p)$

es: Wartezeit auf Kopf bei wiederholtem Münzwurf

* forgetfulness

$X \sim Geom(p)$

$P[X=t]$: $p \cdot (1-p)^{t-1}$

$\mu = E[X]$: $\frac{1}{p}$

$\sigma^2 = Var[X]$: $\frac{1}{p^2} - \frac{1}{p} = \frac{1-p}{p^2}$

$F_X(t) = P[X \le t]$: $1 - (1-p)^t$

Negativbinomial-verteilung

Die Negativbinomialverteilung mit parameter r, p beschreibt den wartezeit auf den r-ten Erfolg bei einer Folge unabhängiger 0-1 Experimente mit erfolgwahrsch. p

$X = \inf\{ t \in \mathbb{N} \mid \sum_{i=1}^{k} Y_i = r \}, \quad Y_i \sim Be(p)$

$NB(1,p) \equiv Geom(p)$

es: Wartezeit auf dritten Kopf beim wiederholten Münzwurf

* simult

$X \sim NB(r,p)$

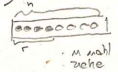
$P[X=t]$: $\binom{t-1}{r-1} \cdot p^r \cdot (1-p)^{t-r}$

$\mu = E[X]$: $\frac{r}{p}$

$\sigma^2 = Var[X]$: $\frac{r \cdot (1-p)}{p^2}$

Hypergeometrische verteilung

Eine Hypergeometrische verteilte ZV beschreibt die anzahl "guter" Kugeln, wenn man aus eine urne mit n Kugeln, davon r "gute", m Mahl ohne Zurücklegen zieht.



es. schweizer lotto

$X \sim HypGeo(n,r,m)$

$P[X=t]$: $\frac{\binom{r}{t} \cdot \binom{n-r}{m-t}}{\binom{n}{m}}$ $t \in \mathcal{W}(x) = \{0,1,\dots, \min(m,r)\}$

$\mu = E[X]$: $\frac{r \cdot m}{n}$

$\sigma^2 = Var[X]$: $\frac{r \cdot m}{n} \cdot \left(1 - \frac{m}{n}\right) \cdot \frac{n-r}{n-1}$

Poisson-verteilung

Die Poisson-verteilung mit parameter $\lambda \in \mathbb{R}^+$ ist eine approximation der Binomialverteilung (# erfolge in 0-1 serie) für kleine p, grossen. $\lambda = np$

Die Approximation ist gut für $np^2 \le 0{,}05$ und damit ein Modell für seltene Ereignisse

es: Anzahl angehende Druckaufträge in n=3600 sekunden, mit einem Jobwahrsch. von $p = 1\%$ pro sekunde, wenn pro sekunde max 1 job entstehen kann

* addit

$X \sim P(\lambda)$

$P[X=t]$: $e^{-\lambda} \cdot \frac{\lambda^t}{t!}$

$\mu = E[X]$: $\lambda = \sum n e^{-\lambda} \frac{\lambda^n}{n!}$

$\sigma^2 = Var[X]$: $\lambda = \sum n^2 e^{-\lambda} \cdot \frac{\lambda^n}{n!} - \lambda^2$

STETIG

Stetige Gleichverteilung

name

description

Bei der stetigen Gleichverteilung oder auch uniformen Verteilung sind alle Werte zwischen zwei Schranken a, b gleichwahrsch.

spezialfall $U(0,1)$

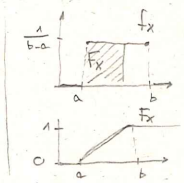
(math)

* interval

es

graph

Δf_X is not the probability, but the change



schreibweise

$X \sim U(a,b)$

$f_X(t)$: $\begin{cases} \frac{1}{b-a} & \text{if } t \in [a,b] \\ 0 & \text{else} \end{cases}$

$\mu = E[X]$: $\frac{a+b}{2}$

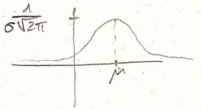
$\sigma^2 = Var[X]$: $\frac{(b-a)^2}{12}$

$F_X(t) = P[X \le t]$: $\begin{cases} 0 & \text{if } t<a \\ \frac{t-a}{b-a} & \text{if } t \in [a,b] \\ 1 & \text{if } t>b \end{cases}$

Normalverteilung

Die Normalverteilung mit Erwartungswert $\mu \in \mathbb{R}$ und varianz $\sigma^2 \in \mathbb{R}^+$ modelliert Summen oder Mittelwerte vieler i.i.d Zufallsvariablen (dies folgt aus dem zentralen Grenzwertsatz) Meistens beschreibt sie die Ergebnisse von Messungen "aus der Natur"

es: gewicht von Kürbissen(zucchi)



$X \sim N(\mu, \sigma^2)$

$f_X(t)$: $\frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(t-\mu)^2}{2\sigma^2}}$

$\mu = E[X]$: μ

$\sigma^2 = Var[X]$: σ^2

$F_X(t) = P[X \le t]$: $\Phi\left(\frac{t-\mu}{\sigma}\right) = P[Y \le \frac{t-\mu}{\sigma}]$

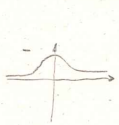
Standard normal-verteilung

$N(0,1)$ heisst standard normal verteilung. Ihre dichte $F_X(t) = \phi(t)$ ist tabelliert.

eigenschaft:

$X \sim N(\mu, \sigma^2) \Rightarrow Y \sim N(0,1)$

$Y = \frac{X-\mu}{\sigma}$



$X \sim N(0,1)$

$f_X(t)$: $\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}$

$\mu = E[X]$: 0

$\sigma^2 = Var[X]$: 1

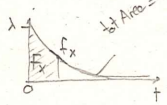
$F_X(t) = P[X \le t]$: tabelliert durch $\Phi(t)$

Exponential-verteilung

(= geometric)

Die exponential verteilung mit parameter $\lambda \in \mathbb{R}^+$ (Ausfallrate) ist das stetige Analogon zur geometrischen verteilung und beschreibt eine Lebensdauer oder eine Wartezeit

* forgetfulness



$X \sim Exp(\lambda)$

$f_X(t)$: $\begin{cases} 0 & \text{if } t<0 \\ \lambda \cdot e^{-\lambda t} & \ge 0 \end{cases}$

$\mu = E[X]$: $\frac{1}{\lambda}$

$\sigma^2 = Var[X]$: $\frac{1}{\lambda^2}$

$F_X(t) = P[X \le t]$: $\begin{cases} 0 & \text{if } t<0 \\ 1 - e^{-\lambda t} & \ge 0 \end{cases}$

Chiquadrat-verteilung (not seen)

Die Chiquadratverteilung mit m Freiheitsgraden taucht (calla fuori) bei beschnitten Tests auf, und entsteht (sosse) durch Quadrieren und Summieren von m standardnormal verteilte ZV

(rarely used to model distribution)

eigenschaft:

$X_m^2 \equiv \sum_{i=1}^{m} X_i^2, \quad X_i \sim N(0,1)$

(square + sum)

$X \sim X_m^2$

t-Verteilung (student-t)

Die t-verteilung mit m freiheitsgraden ist symmetrisch und tritt bei Tests auf.

gamma function: $\Gamma(n) = (n-1)!$

$X \sim t_m$

$f_X(t)$: $\frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{m\pi} \cdot \Gamma\left(\frac{m}{2}\right)} \left(1 + \frac{x^2}{m}\right)^{-\frac{m+1}{2}}$

$\mu = E[X]$: 0 für $n>1$

$\sigma^2 = Var[X]$: $\frac{n}{n-2}$ für $n>2$

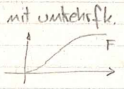
Others:

Pareto-verteilung

Gamma-verteilung

SIMULATION

▷ F eine stetige, streng monoton wachsende, bijektive Verteilungsfunktion mit Umkehrfk. F^{-1}.
▷ $X \sim U(0,1)$ gleichverteilt, $Y = F^{-1}(X)$
Dann hat Y gerade die Verteilungsfunktion F.



Anwendung:
PC liefert immer uniforme ZV zwischen $[0,1]$. Wir wollen aber eine andere Verteilung haben.

Beispiel exponentialverteilung $s = F(t) = 1 - e^{-\lambda t}$
umkehrfunktion $t = F^{-1}(s) = -\frac{\ln(1-s)}{\lambda}$

Sei jetzt x ein zufällig generierter Wert der ZV X, $X \sim U(0,1)$. Es gilt

$$y = F^{-1}(x) = -\frac{\log(1-x)}{\lambda} \sim Exp(\lambda)$$

GEMEINSAME VERTEILUNGEN (joint distrib. funct)

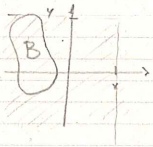
funktionen:
Analog zum fall für eine ZV, existieren für gemeinsame verteilungen

- gemeinsame Gewichtsfunktion (alle X_i diskret) $p(x_1,\dots,x_n) = P[X_1 = x_1, \dots, X_n = x_n]$
- gemeinsame Dichtefunktion (— stetig) $f(x_1,\dots,x_n) =$ —"—
- gemeinsame Verteilungsfunktion $F: \mathbb{R}^n \to [0,1]$
  $(x_1,\dots,x_n) \mapsto F(x_1,\dots,x_n) = P[X_1 \le x_1, \dots, X_n \le x_n]$

→ diskret: $F(x_1,\dots,x_n) = \sum_{t_1 \in W(X_1),\, t_1 \le x_1} \dots \sum_{t_n \in W(X_n),\, t_n \le x_n} P[X_1 = t_1, \dots, X_n = t_n]$

→ stetig: $F(x_1,\dots,x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(t_1,\dots,t_n)\, dt_n \dots dt_1$

darstellung für n=2:
diskret: darstellung als tabelle
stetig: darstellung als Flächengraph



$$P[(X,Y) \in B] = \sum_{\substack{x_i \in W(X) \\ y_i \in W(Y) \\ (x_i,y_i) \in B}} P[X = x_i, Y = y_i]$$

Randverteilungen
Möchte man ausgehend (in words) von einer Gemeinsamen Verteilung bestimmten ZV unbetrachtet lassen und quasi "Einzelnverteilungen" herausfinden, dann:

- DISKRET: Gewichtsfk → unbetrachtete ZV wegsummieren
$$P[X = x] = \sum_{t_1 \in W(X_1)} \dots \sum_{t_{n-1} \in W(X_{n-1})} P[X = x, X_1 = t_1, \dots, X_{n-1} = t_{n-1}]$$
- STETIG: Dichtefunk: → unbetrachtete ZV wegintegrieren
$$f_X(x) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x, y_1, \dots, y_{n-1})\, dy_1 \dots dy_{n-1}$$

Verteilungsfunk: limes bilden! $F_X(x) = \lim_{y_1 \to \infty} \dots \lim_{y_{n-1} \to \infty} F(x, y_1, \dots, y_{n-1})$

⚠ remember to express ranges!

unabhängigkeit von ZV
Die ZV $X_1,\dots,X_n$ heissen unabhängig, falls $F(x_1,\dots,x_n) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n)$

$\Leftrightarrow$ (per teilt i case)

Das ist equivalent zu: diskret: $P[X_1 = x_1, \dots, X_n = x_n] = P[X_1 = x_1] \cdot \dots \cdot P[X_n = x_n]$
stetig: $f(x_1,\dots,x_n) = f_{X_1}(x_1) \cdot \dots \cdot f_{X_n}(x_n)$

i.i.d: sind die ZV gleich verteilt und unabhängig, so heissen sie "iid" = "independent identically distributed"

Bedingte Verteilungen
- **gewichtsfk:** die bedingte gewichtsfunktion einer ZV X, gegeben dass $Y = y$, ist definiert als
$$P[X = x \mid Y = y] = \frac{P[X = x, Y = y]}{P[Y = y]}$$

X, Y sind unabhängig genau dann, wenn $\forall x \in W(X)$:
$$P[X = x \mid Y = y] = P[X = x]$$

- **dichtefk:** die bedingte dichtefunktion einer ZV X, gegeben $Y = y$, ist definiert als
$$f(x \mid y) = \frac{f(x,y)}{f(y)}$$

X, Y unabhängig genau dann, wenn $\forall x$ gilt $f(x \mid y) = f(x)$

GRENZWERTSÄTZE

Schwaches Gesetz der grossen Zahlen – LLN (law of large numbers)
▷ $X_1, X_2, \dots$ unabhängige, aber zumindest paarweise unkorrelierte ZV mit gleichem erwartungswert μ und varianz σ^2. ▷ $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Dann gilt

$$\forall \varepsilon > 0: \lim_{n \to \infty} P[|\overline{X}_n - \mu| > \varepsilon] = 0$$

Starkes Gesetz der grossen Zahlen
▷ $X_1, X_2, \dots$ i.i.d ZV mit endlichem erwartungswert μ. ▷ $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Dann gilt

$$P[\lim_{n \to \infty} \overline{X}_n = \mu] = 1$$ (as number of tests → ∞, value becomes $E[X]$)

Markov ungleichung
▷ X eine ZV, $a \in \mathbb{R}^+$ und $h: \mathbb{R} \to \mathbb{R}^+$ eine fk. Dann gilt

$$P[h(X) \ge a] \le \frac{E[h(X)]}{a}$$

Chebyshev ungleichung
▷ Y eine ZV mit erwartungswert μ, varianz σ^2 (use if no info available on Verteilung)

$$\forall \varepsilon > 0: P[|Y - \mu| \ge \varepsilon] \le \frac{\sigma^2}{\varepsilon^2}$$

Anwendung: sehr leicht berechnbare Schranke für jede beliebige Verteilung. (also not that precise)

Chernoff - schranken
▷ $X_1, \dots, X_n$ i.i.d mit $X_i \sim Be(p)$. ▷ $S_n = \sum_{i=1}^{n} X_i \sim Bin(n,p)$, mit $E[S_n] = np$

$$\forall \varepsilon > 0: P[S_n \ge (1+\varepsilon)E[S_n]] \le \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{(1+\varepsilon)}}\right)^{E[S_n]}$$
(is decreasing)

Anwendung:
Viel einfaches zu rechnen als die verteilungsfunktion der Binomialverteilung, liefert im Gegensatz zur poisson-approximation eine exaktte Schranke

Zentraler Grenzwertsatz
distrib. of X_i:

▷ $X_1, X_2, \dots$ i.i.d. ZV mit $E[X_i] = \mu$, $Var[X_i] = \sigma^2$. ▷ $S_n = \sum_{i=1}^{n} X_i$. Dann gilt

$$\lim_{n \to \infty} P\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \le t\right] = \Phi(t)$$

⚠ $\sigma = \sigma^2$

$$P\left[\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \le t\right] \overset{\text{"}}{\approx} \Phi(t)$$ (for n large)

oft wird die folgende abkürzung benutzt:

$$S_n^* = \frac{S_n - E[S_n]}{\sqrt{Var[S_n]}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$ | $Var[S_n] = \sigma^2 \cdot n$

die entstehende ZV S_n^* heisst Standardisierung von S_n. Der zentrale Grenzwertsatz lässt sich so schreiben als

n gross $\Leftrightarrow S_n^* \overset{approx}{\sim} N(0,1)$
$\Leftrightarrow P[S_n^* \le t] \cong \Phi(t)$
$\Leftrightarrow S_n \overset{approx}{\sim} N(n\mu, n\sigma^2)$

Für $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n} S_n$ gilt
$\overline{X}_n \overset{approx}{\sim} N(\mu, \frac{1}{n}\sigma^2)$

$X_1, \dots, X_n$ i.i.d $\sim N(\mu, \sigma^2) \Rightarrow \overline{X}_n \sim N(\mu, \frac{\sigma^2}{n})$

- $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$
- $\frac{n-1}{\sigma^2} S^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n} (X_i - \overline{X}_n)^2 \sim \chi^2$-verteilt $(n-1)$-freiheitsgraden
- $\overline{X}_n$, S^2 unabhängig
- $\frac{\overline{X}_n - \mu}{S/\sqrt{n}} = \dots = \frac{\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{1}{n-1}\frac{n-1}{\sigma^2}S^2}} \sim t$-verteilt $(n-1)$-freiheitsgr

STATISTIK

Stichproben

In der Statistik möchte man aus beobachteten Daten Rückschlüsse auf die Verteilungen ziehen, welche diese Daten generiert haben.
Eine von einem elementarereignis ω erzeugte Datenmenge heisst Stichprobe (campione d'indagine).
Ihre daten $x_1, \dots, x_n$ heissen realisierungen und werden als werte $X_1(\omega), \dots, X_n(\omega)$ von zufallsvariablen $X_1, \dots, X_n$ betrachtet

→ definiert auf $(\Omega, \mathcal{F}, (P_\vartheta)_{\vartheta \in \Theta})$ $\quad \Theta \subset \mathbb{R}^m$

Schätzer

Betrachtet man jede der zu charakterisierenden zV $X_1, \dots, X_n$ als abhängig von parameter θ_i, so reduziert sich das ziel darauf, diese parameter $\theta_1, \dots, \theta_m \in \Theta$ zu schätzen (stimare)
Ein Schätzer T_i ist eine zV, die in Abhängigkeit der daten genau einen der gesuchten parameter θ_i schätzt.

- **erwartungstreue:** (fedele, unbiased)
Ein Schätzer T_i heisst erwartungstreu für θ_i falls $E[T_i] = \theta_i$
Im mittel, über alle Realisationen schätz er also richtig.
→ $E_\vartheta[T] = \theta$ (verdeckt speaked)
→ $E_\vartheta[T] - \theta =$ "bias"

- **konsistenz:**
Erweitert man einen Schätzer so, dass er weitere Realisierungen derselben zV berücksichtigt (here presente), heisst er konsistent, falls er für eine grosse Anzahl n immer genauer wird.

$$P\left[\lim_{n\to\infty} T_i^{(n)} = \theta_i\right] = 1$$

Maximum-Likelihood-Methode (MLE)

Die Likelyhood-Funktion L gibt die Wahrscheinlichkeit einer Stichprobe unter einem Parameter θ wieder (misura)

$$L(x_1, \dots, x_n; \theta) = \begin{cases} p(x_1, \dots, x_n; \theta) & \to \text{diskret} \\ f(x_1, \dots, x_n; \theta) & \to \text{stetig} \end{cases} \qquad R(x_1, \dots, x_n; \theta_0, \theta_A) := \frac{L(x_1, \dots, x_n; \theta_0)}{L(x_1, \dots, x_n; \theta_A)}$$
→ reject H_0 if R small ($\in [0, c[$)

Sind die zV X_i i.i.d, so gilt

$$L(x_1, \dots, x_n; \theta) = \begin{cases} \prod_{i=1}^n p_X(x_i, \theta) & \to \text{diskret} \\ \prod_{i=1}^n f_X(x_i, \theta) & \to \text{stetig} \end{cases}$$

$\log L(x_1, \dots, x_n; \theta)$ heisst log-likelihood-funktion.
Der Maximum-Likelihood-Schätzer T^{ML} für θ ist dadurch definiert, dass er die Likelihood-funktion L über θ maximiert
Dadurch schätz er θ genau so, dass die gegebene Stichprobe unter θ am wahrscheinlichsten ist.
Da es sich bei L meistens um ein produkt handelt und die log-funkt. streng monoton steigend ↗ ist, kann man auch $\log L$ über θ minimieren, was einfacher zu rechnen ist. • to maximize, derivate

Häufig gebrauchte ML-Schätzer

- $X_i \sim Be(p)$
- $X_i \sim N(\mu, \sigma^2)$

$$T_p = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X}_n$$
$$T_\mu = \text{"} \text{"}$$
$$T_{\sigma^2} = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2 = \left(\frac{1}{n}\sum_{i=1}^n X_i^2\right) - (\overline{X}_n)^2$$

⚠ T_{σ^2} ist nicht erwartungstreu

$$E[T_{\sigma^2}] = \frac{n-1}{n} Var[X]$$

⇒ deshalb benutzt man meistens die empirische Stichprobenvarianz S^2

$$S^2 = T'_{\sigma^2} = \frac{n}{n-1} T_{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$$

TESTS

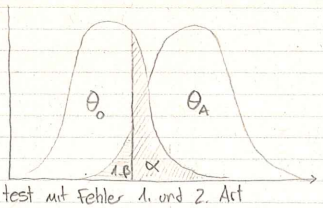
Ausgehend von einer Stichprobe möchte man zwei Verteilungsparameter (bzw. parameterbereiche) gegeneinander abwägen (soppesare):
Eine Nullhypothese Θ_0 und eine Alternativhypothese Θ_A. Man wählt nun eine zV T, namens Teststatistik abhängig von θ, welche die Daten der Stichprobe auswertet (valuta).
Dabei wollen wir die Nullhypothese ablehnen (rifiutare), wenn der realisierte Wert $T(x_1, \dots, x_n)$ in einen Kritischen Bereich oder Verwerfungsbereich K.
Die Entscheidung kann auf zwei Arten falsch herauskommen:

- **fehler 1. Art:** (don't accept right) → die Nullhypothese wird abgelehnt, obwohl sie eigentlich richtig wäre. Die wahrscheinlichkeit für einen Fehler 1. Art heisst Signifikanzniveau α

$$\alpha = P_{\theta_0}[T \in K] \qquad \theta_0 \in \Theta_0$$

- **fehler 2. Art:** (accept wrong) → Die Nullhypotese wird angenommen, obwohl sie falsch wäre. Die wahrsch. für einen fehler 2. Art beträgt $1-\beta$, wobei β als die **Macht** des Tests bezeichnet wird:

$$\beta = P_{\theta_A}[T \in K], \quad \theta_A \in \Theta_A \qquad \text{Fehler 2. Art} = P_{\theta_A}[T \notin K]$$
$$= 1 - P_{\theta_A}[T \notin K]$$



test mit fehler 1. und 2. Art

Da es schwieriger ist, eine Hypotese zu verwerfen, wählen wir als Nullhypotese das gegenteil der eigentlich gewünschten Aussage. (affinare)
Ausserdem wird immer ein Signifikanzniveau α gegeben. Falls sich der Kritische bereich z.B "am rechten rand" von T befindet, berechnet sich der kritische Wert c_α, also dessen untere Grenze (exclusiv), zu:

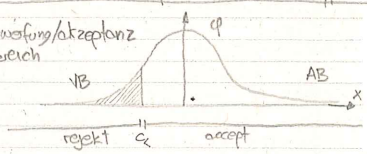
$$P_{\theta_0}[T \geq c] = \alpha \quad \text{(nach c auflösen)}$$
$T < c$

→ it's the smallest significanzniveau α that implies rejecting H_0 ($T(\omega) = t$)

Bei einer zweiseitigen Alternativhypothese sind gegebenenfalls (all'occorrenza, event.) Symmetrien auszunutzen. (sfruttare)
Die Wahrsch. unter der Nullhypothese, den erhaltenen Wert für die Teststatistik zu bekommen heisst **P-wert** Ist er kleiner als α, so verwirft man die Nullhypothese

$p_{rechts} = P[T \geq t \mid H_0]$
$p_{links} \leq$
zweiseitig: $p = 2\min(p_r, p_l)$

↳ the smaller, the more it's against H_0



z-test:

Der z-test ist ein Test für Erwartungswert bei bekannter Varianz σ^2
▷ $X_1, \dots, X_n \sim N(\theta, \sigma^2)$ i.i.d.
Wir wollen die Nullhypothese $\theta = \theta_0$ testen. Teststatistik

$$T = \frac{\overline{X}_n - \theta_0}{\sigma/\sqrt{n}} \sim N(0,1) \qquad \to \text{computed is } T(\omega) = t$$

Den kritischen Bereich K bestimmt man abhängig von der Alternativhypothese Θ_A:

- $\theta_A > \theta_0$ einseitig rechts: $K = [c_>, \infty[$ $\quad c_> = (1-\alpha)\text{-quantil} = \phi^{-1}(1-\alpha)$
- $\theta_A < \theta_0$ einseitig links: $K =]-\infty, c_<]$ $\quad c_< = \alpha\text{-quantil} = \phi^{-1}(\alpha)$
- $\theta_A \neq \theta_0$ zweiseitig: $K =]-\infty, c_{\neq}[\cup]c_{\neq}, \infty[$ $\quad c_{\neq} = (1-\frac{\alpha}{2})\text{-quantil} = \phi^{-1}(1-\frac{\alpha}{2})$
→ verwerfe falls $|t| = \left|\frac{\overline{x}_n - \theta_0}{\sigma/\sqrt{n}}\right| > c_{\neq}$ $\quad (= c_{\neq}(\alpha))$
• use normal (z_{-})

t-test:

Der t-test ist ein Test für Erwartungswert bei unbekannter Varianz σ^2.
▷ $X_1, \dots, X_n \sim N(\theta, \sigma^2)$ und wir möchten die Nullhypothese $\theta = \theta_0$ testen. Teststatistik

$$T = \frac{\overline{X}_n - \theta_0}{\sqrt{S^2/n}} \sim t_{n-1} \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$$

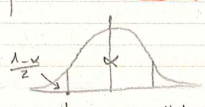
Dabei bezeichnet S^2 den Schätzer für die empirische Stichprobenvarianz.
Der kritische Bereich K lässt sich analog zum z-test mit den Quantilen der t_{n-1}-Verteilung bestimmen.
• use student $(t_{-,-})$

FORMULARIO-STATISTIQUE

	stimed of population	real
moyenne $\overline{X} = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$	$\hat{\mu} = \overline{x}$	μ
variance $V = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2 = \frac{1}{n}\left(\sum_{i=1}^n x_i^2\right) - \overline{x}^2$	$\hat{V} = \frac{n}{n-1} \cdot V$	σ^2
ecart type $s = \sqrt{V}$	$\hat{\sigma} = \sqrt{\hat{V}} = \sqrt{\frac{n}{n-1}} \cdot s$	σ
mediane $\tilde{x}$		

interval de confiance

$$\overline{x} - t_{\frac{1-\alpha}{2}} \cdot \frac{s}{\sqrt{n-1}} \leq \mu \leq \overline{x} + t_{\frac{1-\alpha}{2}} \cdot \frac{s}{\sqrt{n-1}}$$
($\sqrt{n}$ under the s)



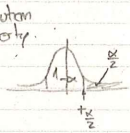
$-t_{\frac{1-\alpha}{2}}$ = Student's t coefficient
• lvl α
• $\nu = n-1$ degrees of freedom

test de student

α = probabilité de se tromper

$$s = \sqrt{\frac{\sum (x_i - \mu)^2}{n}} \qquad t = \frac{\overline{x} - \mu}{s}\sqrt{n-1} \Rightarrow \text{student's distribution } (n-1) \text{ deg. liberty}$$

accept H_0 (depending on α) if $-t_{\frac{\alpha}{2}} \leq t \leq t_{\frac{\alpha}{2}}$



Scaletta:
1. Modell
2. Nullhypoth
3. Alt.-hypoth
4. Teststatistik
5. Verteilung T_{θ_0}
6. Verwerfungsbereich
7. beobachteten wert $T(\omega)$
8. Testentscheid

BEISPIELE

t-test:

Aufgabe: Ein Waschmittelhersteller bringt 5-kg packungen in den Umlauf (in circulation). Die Konsumentenschutzorganisation kauft 25 packungen. Es ergibt sich ein mittel von $\bar{X} = 4{,}9$ kg und eine empirische Stichprobenvarianz $S^2 = 0{,}1$ (kg²).
Die einzelne gewichte seien durch unabhängige $N(\mu, \sigma^2)$ ZV beschrieben.

1) Wie lauten die Hypothesen H_0 und H_A?
2) Wie ist $(\bar{X}-5)/(S/5)$ verteilt unter H_0?
3) Führen sie den t-test auf den 5%-Niveau durch. Wie gross ist der P-wert? Wird die nullhypothese verworfen?
4) Berechnen sie das 95%-Vertrauensintervall für den in 3) beschriebenen test.

lösung:

1) $H_0: \mu = 5$, $H_1: \mu < 5$
2) Verteilung: $\sim t_{S^2-1} = t_{24}$
3) T-statistik: $T_4 = \frac{\bar{X}-\mu_0}{\sqrt{S^2/n}} = -1{,}581$
   P-wert $= P_{\mu_0}[T \le c] = 0{,}0653$.
   Gesucht ist ein c, sodass (?) $P_{\mu_0}[T \le c] = 0{,}05$, bzw $P_{\mu_0}[T > -c] = 0{,}95$.
   Dies ergibt $c = -1{,}711$.
   Since $T_4 = -1{,}581 > c$, wird die Nullhypothese nicht verworfen.
4) Die Formel für das Vertrauensintervall ergibt $I_{0,95} = [4{,}792, 5{,}008]$ (!?)
   Da unser test aber einseitig ist, wird das Intervall $I_{0,95} =]-\infty, 5{,}008]$

Folge von ZV:

Aufgabe: $\sim X_1, \dots, X_n$ iid. Poisson-verteilte ZV mit parameter $\lambda = 2$
Gegeben sei nun die folge $\frac{2n+1}{n^2} \sum_{i=1}^{n} X_i$
wie gross ist der Grenzwert?

lösung: $\frac{2n+1}{n^2} \sum_{i=1}^{n} X_i = \frac{2n+1}{n} \cdot \frac{1}{n} \sum_{i=1}^{n} X_i$

Der zentraler Grenzwertsatz liefert
für $Y \sim N(2, \frac{4}{n})$ $\lim_{n\to\infty} \frac{2n+1}{n^2} \sum_{i=1}^{n} X_i = 2 \cdot E[Y] = 4$

Simulation (1)

Aufgabe: Die ZV X sei Pareto-Verteilt und habe die Verteilungsfunktion:

$$F_X(x|x_0,\alpha) = \begin{cases} 1 - \left(\frac{x}{x_0}\right)^{-\alpha} & \alpha > 1,\ x \ge x_0 \\ 0 & \text{else} \end{cases} \qquad (x \text{ is the variable})$$

$$f_X(x|x_0,\alpha) = \begin{cases} \alpha x_0^{\alpha} x^{-\alpha-1} & \alpha > 1,\ x \ge x_0 \\ 0 & \text{else} \end{cases}$$

Ein auf $[0,1]$ uniform verteiler zufallsgenerator habe nun die Werte 0,237 und 0,733 generiert. Berechne daraus 2 realisierungen der einer Pareto-verteilten ZV mit $\alpha = 2$ und $x_0 = 2.10^6$

lösung:
Die Inverse F_X^{-1} des Verteilungsfunktion, ausgewertet aus den generierten stellen $\in [0,1]$, ergibt die Realisierungen der Verteilung F_X.

für $x \ge x_0$: $F_X^{-1}(y) = \frac{x_0}{(1-y)^{1/\alpha}}$

Jetzt müssen wir nur noch die beiden Werte eingesetz werden.

Simulation (2)

Aufgabe: Die Cauchy-verteilung hat die dichte $f(x) = \frac{1}{\pi \cdot (1+x^2)}$, $x \in \mathbb{R}$
Konstruieren Sie eine Cauchy-verteilte Stichprobe vom Umfang 4 (zora/penuelas) aus der folgenden Expl)
$x_1 = 0{,}417$, $x_2 = 0{,}648$, $x_3 = 0{,}114$, $x_4 = 0{,}665$

lösung:
Die verteilungsfunktion von $X \sim Exp(1)$ ist $F_X(x) = 1 - e^{-x}$.
Wenn man auf exponentialverteilte Werte die Verteilungsfunktion der Exponential-verteilung anwendet, erhält man $[0,1]$-uniform-verteilte Werte.
Dies liefert in diesem bsp durch einsetzen der x: in F die uniform verteilten y_i:
$y_1 = 0{,}341$, $y_2 = 0{,}361$, $y_3 = 0{,}107$, $y_4 = 0{,}486$

Nun muss man (nun = oo/di conseguenza) nur noch die y_i in F^{-1} der Cauchy-Verteilung einsetzen, für die Cauchy-verteilte z_i:

$F(x) = \int_{-\infty}^{x} f(t)\,dt = \frac{1}{2} + \frac{1}{\pi}\arctan(x) \quad \to \quad F^{-1}(y) = \tan\left(\pi\left(y - \frac{1}{2}\right)\right)$

$z_1 = -0{,}544$, $z_2 = -0{,}465$, $z_3 = -2{,}836$, $z_4 = -0{,}043$

Minimum of independent ZV

system of 2 pieces, all fail with prob p on a day. K_i = # days before first fail.
system works only if both work. what is probability system is up for more than n days?
$K = \min(K_1, K_2)$, K_1, K_2 unabh. $P[K > n] = P[K_1 > n \text{ and } K_2 > n] = P[K_1 > n]P[K_2 > n]$
$= (1-p)^{2n} = (1-(2p-p^2))^n \sim Geom(2p-p^2)$

#prim

2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97, ...

partial sum formula $\sum_{i=0}^{n} q^i = \frac{q^{n+1}-1}{q-1}$

infinite sum formula $\sum_{i=0}^{\infty} q^i = \frac{1}{1-q}$ if $|q|<1$

$\sum_{k=j+1}^{\infty} \left(\frac{1}{2}\right)^k = \left(\frac{1}{2}\right)^{j+1} \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k$

exponential series $\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x$

exponential limit $e^x = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n$

binomial formula $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$

FROM EXERCISES

$f_W(t) = \frac{d}{dt} F_Y(t^2) = 2t \cdot f_Y(t^2)$ (by chain so)

faltung die dichte der summe zweier unabhängiger ZV ist gegeben durch die faltung ihrer dichten
$W := X+Y$ $f_W(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx = \int_{-\infty}^{\infty} f_X(z-x) f_Y(x)\,dx$

kettenregel: $(g(f(x)))' = g'(f(x)) \cdot f'(x)$

log: $\log(ab) = \log a + \log b$ $\quad \log(a^b) = b \log a$
$\log\left(\frac{a}{b}\right) = \log a - \log b$

transform: $f_Y = -f_X \implies F_Y = 1 - F_X$

VERI-FALSO

- $X: f_X(x)$, $Y = e^X \Rightarrow f_Y(y) = \frac{f_X(\log y)}{y}$ (see transformation)
- $X \sim N(0,1) \Rightarrow$ correlation $\rho(X, X^2) = 0$
- " , U with $P[U=1] = P[U=-1] = \frac{1}{2}$ unabhängig $\Rightarrow UX \sim N(0,1)$ (reflexo)
- erwartungstreu schätzer, das mittlere quadratisch Schätzfehler immer gleich der varianz der schätzer (if exists)
- $X > 0$, $F_X(x)$, $Y = \frac{1}{X} \Rightarrow F_Y(y) = 1 - F_X\left(\frac{1}{y}\right)$
- $X_1, X_2 \sim N(\mu, \sigma^2)$, $X_3 = 0{,}5(X_1 + X_2) \Rightarrow$ no right answer (you have no info on cov/kohi
- Nonzwurf: P-wert > 30% (einseitig- fairwünze - 4 toss - only 1 head)
- X, Y uncorrelated, $EX = EY = 1$, $\mathrm{Var}X = (\mathrm{Var}Y) = 1$, $E[Y^2] = 2 \Rightarrow \mathrm{Var}[X-Y] = 2$
- $X \sim N(0, \sigma^2)$, $\sigma > 0$. $Z = XY$, Y unabhängig with $P[Y=1] = P[Y=-1] = \frac{1}{2} \Rightarrow X, Z$ unkorrelliert
- to compute P-wert, you need alternative
- $X \sim Bin(3, p)$. $H_0: p_0 = 0{,}5$, $H_A: p_A < 0{,}5$, $\alpha = 5\% \Rightarrow$ no beobachtung rejects H_0
- 1000 phones 300 old, 700 new, pick 200: $P[\#old = 10] = \binom{300}{10}\binom{700}{190} / \binom{1000}{200}$
- $P[X=k] = Ce^{-\lambda} \lambda^k / k!$, $k \in \{1, 2, \dots\} \Rightarrow C = (1-e^{-\lambda})^{-1}$
- $\mathrm{Var}[XY] = (\mathrm{Var}X + E[X]^2)E[Y^2] - E[XY]^2$

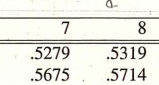
- X_1, X_2 unabh ZV $\sim N(\mu_i, \sigma_i^2)$ $X_3 := \lambda_1 + aX_{14} + bX_2 \sim N(\lambda_1 + a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$
- X_1, X_2 unabh $\sim Exp(\lambda > 0)$. $Y := X_1 + X_2 \sim \Gamma(2, \lambda)$
- $X_1, \dots, X_n$ —"— $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Die verteilung die lässt sich mit verteilung von $\bar{X}_n$ approximieren $\approx N\left(\frac{1}{\lambda}, \frac{1}{n\lambda^2}\right)$

Verteilungsfunktion Φ der Standard normal verteilung

$$\Phi(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz, \quad \text{für } a = 0{,}00 \ldots 3{,}99$$

$\Phi(-a) = 1 - \Phi(a)$

$\Phi^{-1}(a) = -\Phi^{-1}(1-a)$



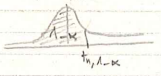
$z_a = \Phi^{-1}(a)$ | $z_{1-\alpha}$

	0	1	2	3	4	5	6	7	8	9
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6408	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.998650	.998694	.998736	.998777	.998817	.998856	.998893	.998930	.998965	.998999
3.1	.999032	.999065	.999096	.999126	.999155	.999184	.999211	.999238	.999264	.999289
3.2	.999313	.999336	.999359	.999381	.999402	.999423	.999443	.999462	.999481	.999499
3.3	.999517	.999534	.999550	.999566	.999581	.999596	.999610	.999624	.999638	.999651
3.4	.999663	.999675	.999687	.999698	.999709	.999720	.999730	.999740	.999749	.999758
3.5	.999767	.999776	.999784	.999792	.999800	.999807	.999815	.999822	.999828	.999835
3.6	.999841	.999847	.999853	.999858	.999864	.999869	.999874	.999879	.999883	.999888
3.7	.999892	.999896	.999900	.999904	.999908	.999912	.999915	.999918	.999922	.999925
3.8	.999928	.999931	.999933	.999936	.999938	.999941	.999943	.999946	.999948	.999950
3.9	.999952	.999954	.999956	.999958	.999959	.999961	.999963	.999964	.999966	.999967

Quantile der t-Verteilung

Ausgewählte Quantile $t_{n,1-\alpha}$ der t-Verteilung

$t_{n,\alpha} = -t_{n,1-\alpha}$



sono già i quantile

n	$t_{0.90}$	$t_{0.95}$	$t_{0.96}$	$t_{0.97}$	$t_{0.975}$	$t_{0.98}$	$t_{0.99}$	$t_{0.995}$
1	3.078	6.314	7.916	10.579	12.706	15.895	31.821	63.657
2	1.886	2.920	3.320	3.896	4.303	4.849	6.965	9.925
3	1.638	2.353	2.605	2.951	3.182	3.482	4.541	5.841
4	1.533	2.132	2.333	2.601	2.776	2.999	3.747	4.604
5	1.476	2.015	2.191	2.422	2.571	2.757	3.365	4.032
6	1.440	1.943	2.104	2.313	2.447	2.612	3.143	3.707
7	1.415	1.895	2.046	2.241	2.365	2.517	2.998	3.499
8	1.397	1.860	2.004	2.189	2.306	2.449	2.896	3.355
9	1.383	1.833	1.973	2.150	2.262	2.398	2.821	3.250
10	1.372	1.812	1.948	2.120	2.228	2.359	2.764	3.169
11	1.363	1.796	1.928	2.096	2.201	2.328	2.718	3.106
12	1.356	1.782	1.912	2.076	2.179	2.303	2.681	3.055
13	1.350	1.771	1.899	2.060	2.160	2.282	2.650	3.012
14	1.345	1.761	1.887	2.046	2.145	2.264	2.624	2.977
15	1.341	1.753	1.878	2.034	2.131	2.249	2.602	2.947
16	1.337	1.746	1.869	2.024	2.120	2.235	2.583	2.921
17	1.333	1.740	1.862	2.015	2.110	2.224	2.567	2.898
18	1.330	1.734	1.855	2.007	2.101	2.214	2.552	2.878
19	1.328	1.729	1.850	2.000	2.093	2.205	2.539	2.861
20	1.325	1.725	1.844	1.994	2.086	2.197	2.528	2.845
30	1.310	1.697	1.812	1.955	2.042	2.147	2.457	2.750
40	1.303	1.684	1.796	1.936	2.021	2.123	2.423	2.704
60	1.296	1.671	1.781	1.917	2.000	2.099	2.390	2.660
120	1.289	1.658	1.766	1.899	1.980	2.076	2.358	2.617

Spezielle unbestimmte Integrale

(Integrationskonstanten weggelassen)

$\int k\,dx = kx, \quad k$ konstant

$\int x^s dx = \frac{1}{s+1}x^{s+1}, \quad s \neq -1$

$\int \frac{1}{x}dx = \log|x|, \quad x \neq 0$

$\int (ax+b)^s dx = \frac{1}{a(s+1)}(ax+b)^{s+1}, \quad s \neq -1$

$\int \frac{1}{ax+b}dx = \frac{1}{a}\log|ax+b|$

$\int \frac{x}{ax+b}dx = \frac{x}{a} - \frac{b}{a^2}\log(ax+b)$

$\int \frac{x}{(a+bx)^2}dx = \frac{b}{a^2\cdot(ax+b)} + \frac{1}{a^2}\log(a \cdot x+b)$

$\int \frac{x}{(ax+b)^n}dx = \frac{1}{a^2}\left(\frac{-1}{(n-2)\cdot(ax+b)^{n-2}} + \frac{b}{(n-1)\cdot(ax+b)^{n-1}}\right), \quad n \neq 1, n \neq 2$

$\int \frac{1}{x^2+a^2}dx = \frac{1}{a}\arctan\frac{x}{a}$

$\int \frac{1}{x^2-a^2}dx = \frac{1}{2a}\log\left|\frac{x-a}{x+a}\right|$

$\int e^{cx}dx = \frac{1}{c}e^{cx}$

$\int a^{cx}dx = \frac{1}{c\cdot\log a}a^{cx}$

$\int xe^{cx}dx = \frac{e^{cx}}{c^2}(cx-1)$

$\int x^2e^{cx}dx = e^{cx}\left(\frac{x^2}{c} - \frac{2x}{c^2} + \frac{2}{c^3}\right)$

$\int \log|x|dx = x(\log|x| - 1)$

$\int \log_a|x|dx = x(\log_a|x| - \log_a e)$

$\int x^s\log x dx = \frac{x^{s+1}}{s+1}\left(\log x - \frac{1}{s+1}\right), \quad s \neq -1, \ x > 0$

$\int \frac{1}{x}\log x dx = \frac{1}{2}(\log x)^2, \quad x > 0$

$\int \sin x dx = -\cos x$

$\int \cos x dx = \sin x$

$\int \tan x dx = -\log|\cos x|$

$\int \cot x dx = \log|\sin x|$

$\int \sin(ax+b)dx = -\frac{1}{a}\cos(ax+b)$

$\int \cos(ax+b)dx = \frac{1}{a}\sin(ax+b)$

$\int \frac{1}{\sin x}dx = \log\left|\tan\frac{x}{2}\right|$

$\int \frac{1}{\cos x}dx = \log\left|\tan\left(\frac{x}{2}+\frac{\pi}{4}\right)\right|$

$\int \sin^2 x dx = \frac{1}{2}(x - \sin x\cos x)$

$\int \cos^2 x dx = \frac{1}{2}(x + \sin x\cos x)$

$\int \tan^2 x dx = \tan x - x$

$\int \cot^2 x dx = -\cot x - x$

log = ln

Tabelle der Binomialkoeffizienten $\binom{n}{k}$

bis $n = 15$

$n\backslash k$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	1													
2	1	2	1												
3	1	3	3	1											
4	1	4	6	4	1										
5	1	5	10	10	5	1									
6	1	6	15	20	15	6	1								
7	1	7	21	35	35	21	7	1							
8	1	8	28	56	70	56	28	8	1						
9	1	9	36	84	126	126	84	36	9	1					
10	1	10	45	120	210	252	210	120	45	10	1				
11	1	11	55	165	330	462	462	330	165	55	11	1			
12	1	12	66	220	495	792	924	792	495	220	66	12	1		
13	1	13	78	286	715	1287	1716	1716	1287	715	286	78	13	1	
14	1	14	91	364	1001	2002	3003	3432	3003	2002	1001	364	91	14	1
15	1	15	105	455	1365	3003	5005	6435	6435	5005	3003	1365	455	105	15